World Scientific Research Journal Volume 11 Issue 3, 2025
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0004

Maize Seedling and Weed Detection Using BFSL-YOLOv8

Jia Su[1], Ziqi Song[2], Xuejing Wang[3], Jiang Ma[4, *]

[1]School of Mechanical Engineeing, North China University of Science and Technology, Tangshan 063210, China

[2]College of Foreign Languages, North China University of Science and Technology, Tangshan 063210, China

[3]School of Economics and Management, North China University of Science and Technology, Tangshan 063210, China

[4]School of Science, North China University of Science and Technology, Tangshan 063210, China

* Corresponding author

Abstract

Addressing the urgent need for weed control in maize seedlings due to the low detection accuracy in complex agricultural environments, this paper proposes a BFSL-YOLOv8 model. paper proposes a BFSL-YOLOv8 model. The model enhances multi-scale feature extraction and fusion by integrating a Spatial Pyramid Pooling with Large Separable Kernel Attention and skip connections (SPPF_LSKA) module into the neck of YOLOv8 and efficiently captures long-range dependencies by introducing an improved BiFormer module in the head. Experiments were conducted on a publicly available real-world field monitoring dataset under Experiments were conducted on a publicly available real-world field monitoring dataset under specific hardware and software environments, using stochastic gradient descent (SGD) for training. achieves a mean Average Precision (mAP) of 99.5% for maize detection and improves the mAP at an Intersection over Union (IoU) threshold of 0.5 (mAP50) for weed detection by 0.5 percentage points to 64.8% compared to the baseline YOLOv8. The model has 3.70M parameters and a processing time of 4.4ms per image. F1-confidence curve analysis indicates that BFSL-YOLOv8 exhibits good robustness in complex scenarios, with an optimal confidence threshold of 0.292 and an average confidence threshold of 1.5 times. of 0.292 and an average F1-score of 0.82 for all classes. The experiments validate the effectiveness of the proposed method, providing a new solution for accurate and efficient weed detection in maize. accurate and efficient weed detection in maize fields for precision agriculture.

Keywords

Maize, Weed Detection, YOLOv8, BiFormer, SPPF, LSKA.

1. INTRODUCTION

In China, maize, as one of the major food crops, plays a crucial role in the stability of agricultural production. According to research, maize is susceptible to weed infestation during the growth process, especially in the seedling stage[1-2], which poses a great challenge to the yield and quality of maize. Currently, weed control in agricultural fields mainly relies on chemical herbicides, which is not only time-consuming and laborious, but also leads to the enhancement of weed resistance and causes serious impacts on the environment such as soil pollution and ecological imbalance[3-8]. Therefore, the development of modern, precise, and

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0004

environmentally friendly weed detection and control technologies has become an urgent need for sustainable agricultural development.

In recent years, deep learning technology has been widely used in agricultural automated detection[9]. Can Wang et al[10] proposed a corn weed recognition algorithm based on convolutional neural network to extract multiscale hierarchical features, combining multiscale features with hyperpixel segmentation. Fan Xiangpeng et al.[11] proposed a Faster R-CNN-based algorithm for recognizing seedlings and associated weeds in cotton fields. Le Zhang et al.[12] proposed a VGG-16 based Faster R-CNN deep network for oilseed rape recognition, which utilizes RPN to generate high-quality region suggestion frames, and achieves convolutional layer sharing with VGG-16 feature extraction network, with a large model and accuracy to be improved. Jie Kang et al.[13] Introduced the idea of multi-scale fusion into the SSD model to recognize sugar beet and weeds, and the accuracy and speed of recognition were effective. Song Peng et al.[14] Proposed multi-channel larval shoots localization method based on Cascade R-CNN by combining RGB and depth images, with an average accuracy of 97.5% and a fast processing speed. Jialin Yu et al[15] Detecting dogbane lawn and its accompanying weeds using deep convolutional neural network containing AlexNet, GoogleNet, and VGGNet, after comparison, VGGNet has the best detection effect, but the model complexity is high, which is difficult to satisfy the demand of lightweight and high efficiency of field weeding. Jun Sun et al.[16] proposed a depth-separable model combined with multi-channel residual blocks to achieve fast and accurate detection of sugar beet and weeds. In addition, Qingkuan Meng et al.[17] proposed an SSD model dedicated to weed detection in corn fields by combining a lightweight model with a feature layer fusion mechanism. Although the lightweight design effectively reduces the computational complexity, it affects the detection accuracy. Chien-Yao Wang et al.[18] Lightweighting YOLOv4 and designing a YOLOv4-tiny model, the speed is significantly improved when tested on GPU. Yuqing Yang et al.[19] proposed a weed detection algorithm for peanut seedlings based on YOLOv4-Tiny-CBAM, which has an average weed detection accuracy of 90.39%, and a single detection time of 44.79 ms, compared with the R-CNN algorithm and SSD algorithm, which have a higher detection rate. Wenli et al.[20] proposed an algorithm based on YOLOv4-Tiny-CBAM, which combines the MSRCR algorithm and PP-LCNet network proposed a lightweight weed recognition model based on YOLOv5[21]. However, these existing methods still have some limitations in the efficiency and accuracy of weed detection. Although the lightweight model can improve the operation speed, it may sacrifice the detection accuracy; although the complex model has higher accuracy, the computational complexity is large, which is not conducive to practical applications.

Therefore, this study proposes a BFSL-YOLOv8 model that combines an improved BiFormer[22] module and the SPPF_LSKA[23] jumping link that incorporates large kernel attention. The BiFormer module, leveraging its distinctive architecture and attention mechanism, facilitates more efficient processing of image features, thereby enhancing the model's target detection capabilities, particularly in complex field environments, proving effective for seedling corn and weed detection. The SPPF_LSKA module enhances small target detection through multi-scale pooling and a separable large kernel attention mechanism. Furthermore, the incorporation of skip connections promotes the retention of low-level feature information, enabling effective fusion with high-level features to further improve overall model performance, providing a more efficient and accurate solution for agricultural weed detection.

2. EXPERIMENTAL DATA

The two-dimensional image dataset used in this study was compiled from publicly available online resources. The dataset comprises two target classes: corn and weeds, with a total of 8,500 images, partitioned into a training set of 7,000 images and a test set of 1,500 images. To

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0004

enhance the dataset size and improve model generalization, a series of image augmentation techniques were applied, including blurring, contrast adjustment, and color space transformations. These augmentations enable the model to better adapt to variations in image quality and lighting conditions.

Figure 1. Data enhancement

3. MODELING

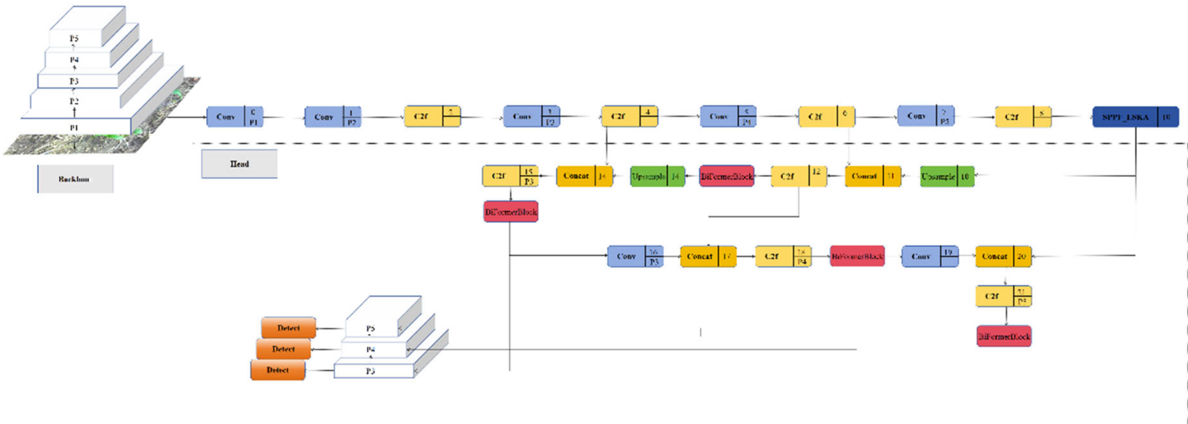

Figure 2. The network architecture diagram of BFSL-YOLOv8 model.

This study aims to improve the target detection performance of the YOLOv8 model in complex agricultural scenarios by proposing a new SPPF_LSKA jump connection module and introducing the BiFormer module to address its shortcomings in multi-scale feature extraction and enhancement. This section will first briefly review the basic architecture of YOLOv8, and then focus on the innovative module proposed in this paper and its application

(1) YOLOv8 model overview

YOLOv8 is a state-of-the-art single-stage target detection model that uses an end-to-end framework to directly predict the bounding boxes and categories of targets in an image. The model consists of three main components: backbone network (Backbone), neck network (Neck) and prediction head (Head).

Backbone network: used to extract the features of the input image. Let the input image be I. After processing by the backbone network, the output feature map is denoted as F:

$$F = \text{Backbone}(I) \tag{1}$$

Neck: YOLOv8 implements multi-scale feature fusion via a modified Path Aggregation Network (PAN-FPN) architecture. This architecture combines top-down and bottom-up pathway aggregation and integrates a Spatial Pyramid Pooling Fast (SPPF) module to refine

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0004

feature representations.Prediction Head (Head): generates the bounding box b_i and the category probability c_i of the target based on the fused feature map F:

$$y = \text{Head}(F) = \{(b_i, c_i) \mid i \in \{1, 2, \ldots, N\}\} \tag{2}$$

where b_i denotes the coordinates of the i-th prediction frame, c_i denotes the corresponding category probability, and N is the number of prediction frames.

(2) SPPF_LSKA jump connection

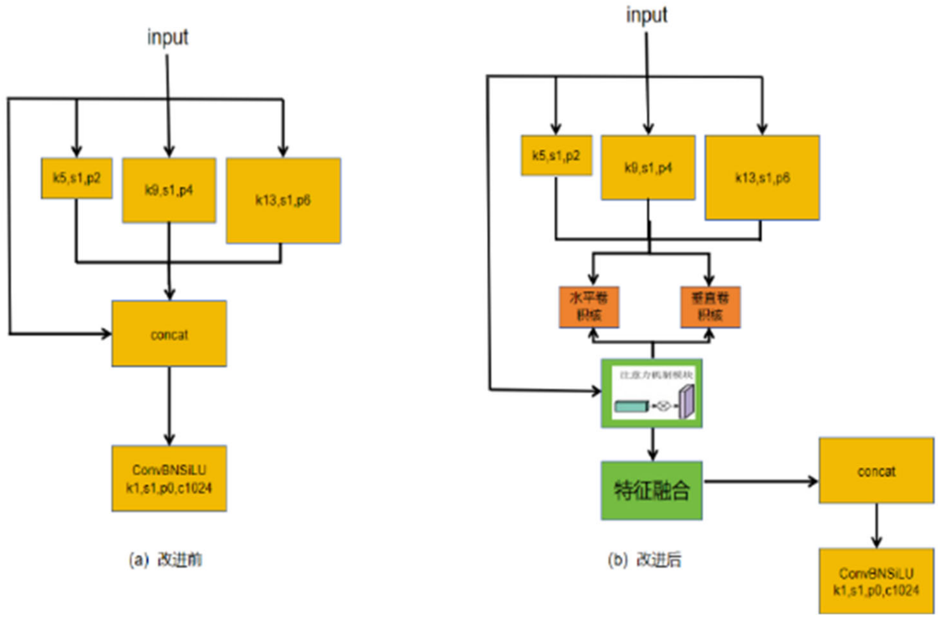

Figure 3. Comparison of SPPF Performance

In order to enhance the multi-scale feature extraction and feature enhancement capability of the model, a new SPPF_LSKA jump connection module is proposed in this study. This module organically combines SPPF, Large Separable Kernel Attention (LSKA) and jump connection mechanism, aiming at extracting and fusing multi-scale features more efficiently and retaining the detailed information of low-level features. As shown in Figure 2, the SPPF_LSKA structure replaces the original SPPF module of YOLOv8.

Given the feature map F output from the backbone network, multi-scale context information is extracted using SPPF. SPPF performs maximum pooling operations in parallel with multiple pooling kernels of different sizes (k_i) and splices the results. Subsequently, the output of the SPPF is passed to the LSKA module for feature enhancement. LSKA first processes the input feature map F using the horizontal convolution kernel (Conv_v) and the vertical convolution kernel (Conv_h) to obtain the attention feature map A. Then, the attention map A, which has been transformed by the 1×1 convolution, is multiplied element-by-element with the input feature map F (⊗) to generate the enhanced feature F', and finally, the output F' of LSKA is element-wise summed with the original input F through jump-joining to obtain the final fused feature map F_{fusion}.

$$A = \text{Conv}_v(\text{Conv}_h(F)) \tag{3}$$

$$F = F \otimes \text{Conv}_{1\times1}(A) \tag{4}$$

$$F_{fusion} = F + F' \tag{5}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0004

where n denotes the number of pooling layers, k_i denotes the kernel size of the i-th pooling layer, and $\otimes$ denotes the element-by-element product.

(3) Application of the BiFormer module to predictor heads

In order to further improve the model's detection performance for multi-scale targets in complex agricultural scenarios, this study introduces the BiFormer module after the P3, P4, and P5 outputs of the YOLOv8 prediction head (Head), respectively. At the core of the BiFormer module is the Bi-level Routing Attention (BRA) mechanism which aims to capture long-range dependencies in feature graphs in an efficient way. Compared with the traditional multi-head self-attention mechanism, BRA selects the most relevant key-value pairs for computation in a content-aware way by introducing coarse-grained region-level routing and fine-grained token-to-token attention, effectively avoiding the redundancy brought by global computation and significantly reducing the computational complexity, while maintaining a strong feature representation capability. This mechanism is particularly suitable for agricultural scenarios with large target size differences and complex backgrounds, such as seedling corn and weed detection. For a detailed structure of the BiFormer module and a complete description of the BRA mechanism, please refer to the related literature.

4. EXPERIMENTAL DESIGN

(1) Experimental environment

The experiments were performed on a computer with Windows 11 operating system, Intel Core i7-13620H processor, 16GB RAM and NVIDIA RTX 3090 GPU. The algorithms are developed and implemented based on the PyTorch deep learning framework and GPU training using CUDA acceleration.

(2) Hyperparameter setting

This study uses the Ultralytics YOLOv8 framework for target detection in corn and weeds. Model training was performed using a stochastic gradient descent (SGD) optimizer with momentum, which was automatically selected for the framework. The key training hyperparameters are set as follows: initial learning rate is 0.01, final learning rate is 0.01, momentum is 0.937, and weight decay is 0.0005. the training period is 100, and the batch size is 8. the warmup strategy is used for the first 3 epochs. The weights of each part of the loss function are 7.5 for bounding box loss, 0.5 for classification loss, and 1.5 for DFL loss.

(3) Assessment indicators

The following metrics were used in this study to evaluate the model performance. Mean Average Precision (mAP) is the most commonly used evaluation metric in target detection tasks to measure the average detection accuracy of the model over all categories. In this work, the mAP value at the Intersection over Union (IoU) threshold of 0.5 is used, denoted as mAP50. IoU is defined as the ratio of the intersection area of the Bounding Box and the Ground Truth Box to the area of the Union area. To evaluate the performance of the model under different confidence thresholds and to select the optimal confidence threshold, this study uses the F1-score and the F1-Confidence Curve. The F1-score is the harmonic mean of Precision and Recall, which is a combined measure of the accuracy and completeness of detection, and is defined in Equation (8). Among them, Precision is defined in equation (6) , and Recall is defined in equation (7).

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

World Scientific Research Journal Volume 11 Issue 3, 2025
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0004

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{8}$$

Where TP (True Positive) denotes true cases, i.e., the number of samples that were correctly predicted as positive cases; FP (False Positive) denotes false positive cases, i.e., the number of samples in which negative cases were incorrectly predicted as positive cases; and FN (False Negative) denotes false negative cases, i.e., the number of samples in which positive cases were incorrectly predicted as negative cases.

5. EXPERIMENTAL RESULTS AND ANALYSIS

(1) Comparative experiments

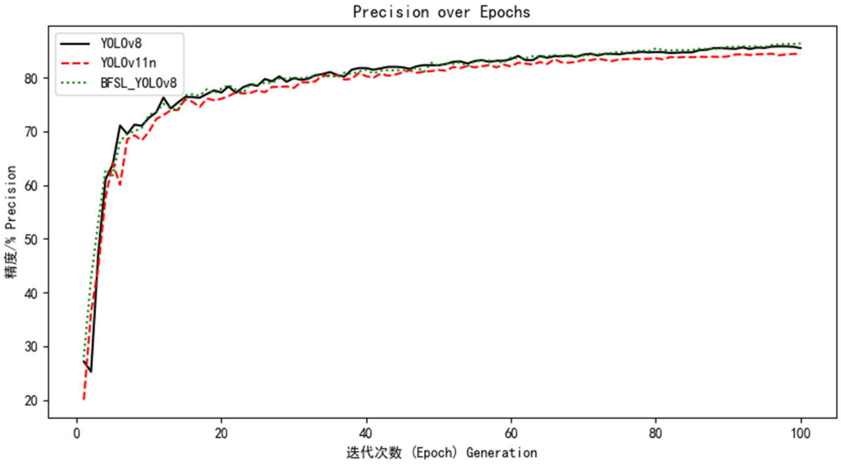

Figure 4. Comparison of training accuracy of different models

Figure 4 presents the training accuracy curves of YOLOv8, YOLOv11n, and BFSL-YOLOv8. As shown in the figure, YOLOv11n demonstrates faster convergence during the early training phase; however, its accuracy plateaus and even declines slightly in the later stages. In contrast, the accuracy of YOLOv8 and BFSL-YOLOv8 continues to improve steadily and eventually surpasses that of YOLOv11n. This observation suggests that YOLOv11n may be susceptible to overfitting or insufficient training, while the training process of YOLOv8 and BFSL-YOLOv8 exhibits greater stability.

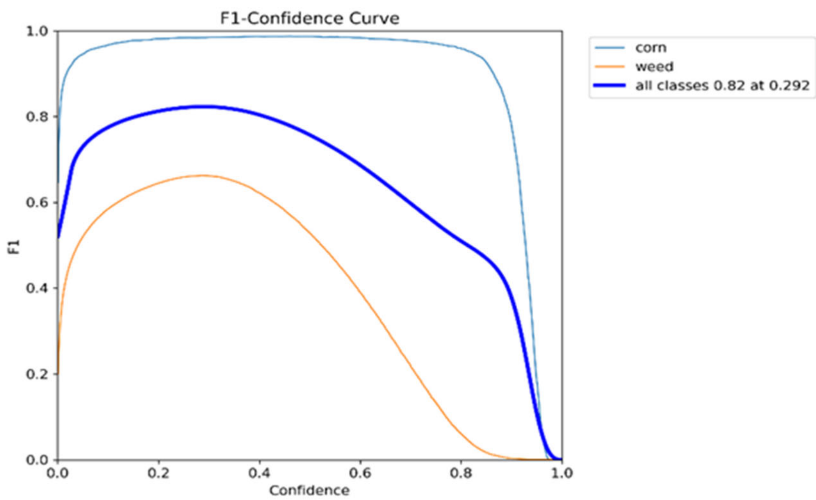

Figure 5. F1-Confidence Curve for BFSL-YOLOv8

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0004

Table 1. Effect of different methods on model performance

model	mAP50(%)	Corn mAP50(%)	Weed mAP50(%)	Parameters (M)	Inference time (ms)	Optimal Threshold
YOLOv8	81.9	99.4	64.3	3.04	2.5	0.296
YOLOv8+BiFormer	81.9	99.4	64.4	3.46	2.3	0.296
YOLOv8+SPPF_LSKA	82.0	99.4	65.0	3.28	4.1	0.293
BFSL-YOLOv8	82.1	99.5	64.8	3.70	4.4	0.292

Table 1 summarizes the effect of different methods on model performance. Observing Table 1, it can be seen that the performance of the models in terms of average detection accuracy is relatively close, with fluctuations ranging from 81.9% to 82.1%. Among them, the BFSL-YOLOv8 model achieves the highest mAP50 of 82.1%, which is improved by 0.2 percentage points, or 0.24%, compared with the baseline model, YOLOv8. YOLOv8+SPPF_LSKA follows with a mAP50 of 82.0%, which is also slightly higher than the baseline model. Although the overall improvement is relatively limited, it indicates that BFSL-YOLOv8 has a slight advantage in overall detection performance. All models performed well on maize detection, so the improvement had no significant effect on maize detection.

This study focuses on the improvement of weed detection performance.YOLOv8 + SPPF_LSKA performed the best in this aspect. The weed mAP50 of BFSL-YOLOv8 was 64.8%, which was slightly lower than YOLOv8 + SPPF_LSKA, but significantly better than the baseline model YOLOv8, with an enhancement of 0.5 percentage points, as well as compared to YOLOv8+BiFormer, with a 0.4 percentage point improvement. This validates the effectiveness of the improved strategy of BFSL-YOLOv8 for weed detection.

BiFormer's main contribution is a reduction in inference time and a slight increase in the number of parameters, suggesting that it optimizes computational efficiency with essentially no loss of accuracy. SPPF_LSKA significantly improves weed detection accuracy but also significantly increases inference time and the number of parameters. BFSL-YOLOv8 combines the improvements of the two and achieves some improvement in weed detection.

Figure 5 shows that the average F1 score across all categories peaked at 0.82 when the confidence threshold was approximately 0.292, which matches the optimal confidence threshold for this model in Table 1, verifying the consistency of the results. Figure 5 also reflects that the model has significantly higher F1 scores on the corn assay than on the weeds, consistent with the results in Table 1 where corn mAP50 is much higher than weed mAP50.

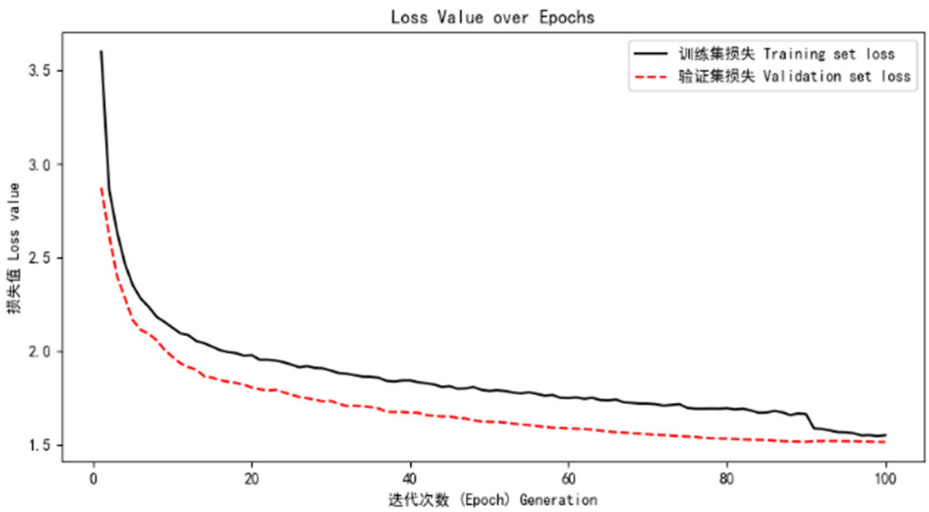

Figure 6. Loss variation curves of BFSL-YOLOv8

Figure 6 shows the loss curves during the training process for each model. As the number of training epochs increases, both the training and validation losses gradually decrease. The small difference between the training and validation losses suggests proper model training and the absence of significant overfitting. This observation aligns with the smooth trend of the accuracy curves shown in Fig.

6. CONCLUSION

In this study, YOLOv8 served as the baseline model, and the BFSL-YOLOv8 model was proposed with a focus on enhancing weed detection performance. Experimental results demonstrate that BFSL-YOLOv8 exhibits a marginal improvement in weed detection compared to the baseline, validating the efficacy of the proposed method. While BFSL-YOLOv8 incurs a slight increase in both the number of parameters and inference time, this trade-off is deemed acceptable in certain application contexts given its potential benefits in weed detection accuracy. Future research will investigate BFSL-YOLOv8's performance with respect to other criteria, such as robustness and generalization capability, conduct a more in-depth analysis of its advantages in weed detection, and explore strategies to further reduce the model's computational complexity.

REFERENCES

[1] Li Shaopeng, et al. Progress of research on weed damage and comprehensive prevention and control technology in corn fields [J]. Corn Science, 2020, 28(4): 188-194.

[2] KNEZEVICH S R, HORAK M J, PATRICK K A. Corn (Zea mays L.) yield loss assessments from common sunflower (Helianthus annuus L.) [J]. Weed Science, 1997, 45(2): 202-207.

[3] WILSON R G, WESTERMAN P W. Weed thresholds for postemergence herbicide applications in corn (Zea mays) [J]. Weed Technology, 1991, 5(2): 392-398.

[4] HEAP I. Global perspective of herbicide-resistant weeds [J]. Pest Management Science, 2014, 70(9): 1306-1315.

[5] SILVA V, et al. Pesticide fate and transport in agricultural soils: an overview [J]. Science of the Total Environment, 2018, 612: 1322-1335.

[6] W. B. Wu, et al. Research progress on the effects of herbicides on soil environment [J]. Journal of Agricultural Environmental Science, 2019, 38(1): 1-10.

[7] FENNIMORE S A, et al. Environmental impacts of pesticides used in agriculture [J]. Advances in Agronomy, 2018, 150: 1-60.

[8] MOSTAFALOU S, ABDOLLAHI M. Pesticides and human chronic diseases: evidence-based toxicology [J]. Toxicology and applied pharmacology, 2017, 317: 1-12.

[9] KAMILARIS A, PRENAFETA-BOLDÚ F. Deep learning in agriculture: a survey [J]. Computers and Electronics in Agriculture, 2018, 147: 70-90.

[10] Can Wang, Xinhui Wu, Zhiwei Li. Multi-scale hierarchical feature extraction based on convolutional neural network to recognize corn weeds[J]. Journal of Agricultural Engineering, 2018, 34(5): 144-151.

[11] FAN Xiangpeng, ZHOU Jianping, XU Yan, et al. Optimized Faster R-CNN-based weed identification and localization in cotton seedling stage[J]. Journal of Agricultural Machinery, 2021, 52(5): 26-34.

[12] Zhang Le, Jin Xiu, Fu Leiyang, et al. A weed identification method for oilseed rape field based on Faster R-CNN deep network[J]. Advances in Lasers and Optoelectronics, 2020, 57(2): 304-312.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202503_11(3).0004

[13] Jie Xu, Gang Liu, Guo Guofa. Weed detection method based on multi-scale fusion module and feature enhancement[J]. Journal of Agricultural Machinery, 2022, 53(4): 254-260.

[14] [14]PENG S, KEYI C, LONGFU Z, et al. An improved cascade R-CNN and RGB-D camera-based method for dynamic cotton top bud recognition and localization in the field[J]. Computers and Electronics in Agriculture, 2022, 202.

[15] YU J, SCHUMANN W A, SHARPE M S, et al. Detection of grassy weeds in bermudagrass with deep convolutional neural networks[J]. Weed Science, 2020, 68(5): 1-31.

[16] J. Sun, W. Tan, S. Wu, et al. Multi-channel depth-separable convolutional modeling for real-time recognition of sugar beet and weeds in complex background[J]. Journal of Agricultural Engineering, 2019, 35(12): 184-190.

[17] MENG Qingkuan, ZHANG Man, YANG Xiaoxia, et al. Corn seedling and weed recognition based on lightweight convolution combined with feature information fusion[J]. Journal of Agricultural Machinery, 2020, 51(12): 238-245.

[18] WANG C Y, BOCHKOVSKIY A, LIAO H. Scaled-YOLOv4: Scaling cross stage partial network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13029-13038.

[19] YANG Yuqing. Research and application of peanut seedling stage weed identification based on YOLOv4-Tiny-CBAM [D]. Anhui Agricultural University, 2022. (DOI information is usually not placed in the references, and dissertations need to include the name of the school)

[20] JI Wenli, LIU Zhou, XING Haihua. A lightweight method for weed identification in farmland based on YOLO v5[J]. Journal of Agricultural Machinery, 2024, 55(1): 212-222.

[21] Ultralytics.yolov5[EB/OL].https://github.com/ultralytics/yolov5 , 2024-10-26.

[22] HE J, ZHANG S, YANG C, et al. Pest recognition in microstate state: an improvement of YOLOv7 based on Spatial and Channel Reconstruction Convolution for feature redundancy and vision transformer with Bi-Level Routing Attention[J]. Frontiers in Plant Science, 2024, 15: 1327237.

[23] WAI K L, LAI-MAN P, UR A Y R. Large Separable Kernel Attention: Rethinking the Large Kernel Attention design in CNN[J]. Expert Systems With Applications, 2024, 236.